

PCT/KR 03/02322
RO/KR 31.10.2003

REC'D 25 NOV 2003

WIPO

PCT



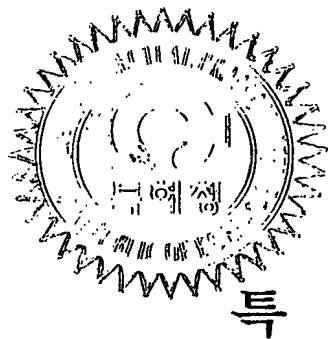
별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출원 번호 : 10-2002-0067416
Application Number

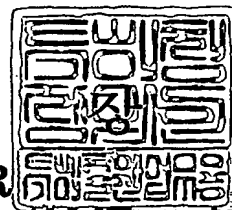
출원 년 월 일 : 2002년 11월 01일
Date of Application NOV 01, 2002

출원인 : 한국전자통신연구원
Applicant(s) Electronics and Telecommunications Research Ins



2003 년 10 월 31 일

특 허 청
COMMISSIONER



PRIORITY DOCUMENT
REMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0001
【제출일자】	2002.11.01
【발명의 명칭】	소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템 및 방법
【발명의 영문명칭】	web contents transcoding system and method for small display devices
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【성명】	권태복
【대리인코드】	9-2001-000347-1
【포괄위임등록번호】	2001-057650-1
【대리인】	
【성명】	이화익
【대리인코드】	9-1998-000417-9
【포괄위임등록번호】	1999-021997-1
【발명자】	
【성명의 국문표기】	신희숙
【성명의 영문표기】	SHIN,Hee Sook
【주민등록번호】	750305-2690217
【우편번호】	302-150
【주소】	대전광역시 서구 만년동 158번지 만년청사빌딩 405호
【국적】	KR
【발명자】	
【성명의 국문표기】	이동우
【성명의 영문표기】	LEE,Dong Woo
【주민등록번호】	721215-1785426
【우편번호】	302-283
【주소】	대전광역시 서구 월평3동 누리아파트 109동 904호
【국적】	KR

【발명자】

【성명의 국문표기】

마평수

【성명의 영문표기】

MAH, Pyeong Soo

【주민등록번호】

610829-1481310

【우편번호】

305-340

【주소】

대전광역시 유성구 도룡동 타운하우스 3-201

【국적】

KR

【발명자】

【성명의 국문표기】

김범호

【성명의 영문표기】

KIM, Bum Ho

【주민등록번호】

740621-1074519

【우편번호】

137-063

【주소】

서울특별시 서초구 방배3동 541-3 명빌라 2동 102호

【국적】

KR

【심사청구】

청구

【취지】

특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인
권태복 (인) 대리인
이화익 (인)

【수수료】

【기본출원료】

20 면 29,000 원

【가산출원료】

19 면 19,000 원

【우선권주장료】

0 건 0 원

【심사청구료】

6 항 301,000 원

【합계】

349,000 원

【감면사유】

정부출연연구기관

【감면후 수수료】

174,500 원

【기술이전】

【기술양도】

희망

【실시권 허여】

희망

【기술지도】

희망

【첨부서류】

1. 요약서·명세서(도면)_1통

【요약서】**【요약】**

본 발명은 소형 화면의 단말을 가진 사용자가 인터넷에 접속하여 웹 서비스를 사용하고 자 할 경우, 기존의 일반 데스크탑 PC의 디스플레이 성능에 적합하도록 작성된 웹 문서를 소형 화면에서도 효율적으로 표현되어질 수 있도록 변환해 주기 위한 시스템 및 방법에 관한 것이다.

본 발명은 웹 문서의 시각적 표현을 주도하는 태그 정보에 대한 분석과 내용 단위로 분리된 조각들의 적절한 재구성 및 인덱스 생성을 통하여 소형 단말에서도 편리한 인터페이스를 사용하여 웹 문서를 브라우징할 수 있도록 한다. 특히 기존의 변환 방법들이 웹 문서 내용의 축약 및 삭제를 통해서 소형 단말을 지원한 것에 반하여, 본 발명은 원본 문서의 내용을 최대한 반영하고자 변경을 최소화한 반면 웹 문서의 적절한 재구성 및 재표현을 통하여 단말의 디스플레이 성능에 맞는 문서를 제공한다. 또한 웹 문서의 분석은 구조적 의미를 가지는 태그와 그 속성값을 주로 이용하고, 유사 내용을 컴포넌트 블록 단위로 묶어서 블록 단위의 재배치를 시도하고 인덱스 부분을 추출하여 재표현하며 특정 텍스트 위주의 내용 단위에 대해 음성이 지원되는 마크업 언어로 변환하여 제공한다.

【대표도】

도 4

【색인어】

웹 서비스, 콘텐츠 변환, 소형 단말기, 디스플레이폭, 인덱스, 컴포넌트

【명세서】

【발명의 명칭】

소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템 및 방법 { web contents transcoding system and method for small display devices }

【도면의 간단한 설명】

도 1은 시각적인 분리 및 그룹화로 상이한 내용 블록을 표현하는 웹 문서의 예시도,

도 2는 본 발명에 따른 소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템의 모듈 구성 개념도,

도 3은 테이블 태그의 표현 계층 관계도,

도 4은 본 발명에 따른 소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템의 동작 과정 순서도,

도 5는 도 4에 도시된 웹 문서 분석 단계의 세부 알고리즘 순서도,

도 6은 도 4에 도시된 컴포넌트 블록 설정 과정의 세부 알고리즘 순서도,

도 7a,7b는 본 발명에 따른 웹 문서의 분석 및 컴포넌트 블록 추출 과정 설명을 위한 예시도,

도 8은 도 4에 도시된 컴포넌트 블록 분류 과정의 세부 알고리즘 순서도,

도 9a,9b는 본 발명에 따른 웹 콘텐츠 변환 결과의 예시도.

도면의 주요부분에 대한 부호의 설명

101: 컴포넌트 102: 컴포넌트 블록

201: 전처리기 202: 클라이언트 프로파일 분석기
203: 구조분석기 204: 이미지 컨버터
205: 컴포넌트 블록 추출기 206: 컴포넌트 블록 분류기
207: 인덱스 생성기 208: 음성 마크업 생성기
209: 커스텀화된 HTML 생성기

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

- <17> 본 발명은 웹 콘텐츠 변환기술에 관한 것으로, 더욱 상세하게는 일반 데스크탑 PC의 디스플레이 성능에 적합하도록 작성된 웹 문서를 소형 화면에서도 효율적으로 표현되어질 수 있도록 변환해 주는 소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템 및 방법에 관한 것이다.
- <18> 최근들어, 이동 통신 기술과 소형 단말 기술의 발전이 가속화되면서, 이 기술들과 인터넷의 접목이 무선 인터넷 환경을 만들어 냈고, 언제 어디서든지 웹을 사용하고자 하는 사람들의 욕구를 충족시켜주기 시작하였다. 하지만 유선 인터넷상의 수 많은 웹 정보는 데스크탑 컴퓨터의 화면 크기에 맞게 제작되어졌고, 이를 소형 디스플레이를 가지는 단말을 통하여 브라우징할 경우, 단말의 성능을 초과하는 콘텐츠 정보로 인하여 단말에서 제대로 표현하지 못하는 문제점이 발생된다.
- <19> 이러한 문제를 해결하기 위해 다양한 콘텐츠 변환기법이 제안되었으나, 초기 셀룰러 폰 (Cellular Phone) 계열의 단말 또는 저성능의 PDA 등을 지원하기 위한 기법들은 단순한 텍스트

요약으로의 변환이 주가 되었기 때문에 사용자가 요구하는 많은 정보를 제대로 표현하지 못했다. 이는 단말 성능의 제한과 텍스트 또는 HDML, WML 등의 간단한 표현 능력을 가진 무선 인터넷 마크업 언어를 주로 사용했기 때문이다.

<20> 이러한 변환은 기존 웹 정보의 일부분만을 추출하여 변환하므로 복잡한 구조에 많은 이미지와 정보를 한꺼번에 표현하는 현재의 웹 페이지를 정확하게 변환하기는 어려운 문제점이 있다.

<21> 이후 고성능의 PDA, 핸드 헬드(Hand-Held) PC 등의 단말이 등장하면서 이를 위한 변환기법이 계속 연구되었고, 그 결과의 IBM의 WebSphere Transcoding Publisher, Synglass Prism 등의 서버에서 동작하는 변환 툴이 나타났다. 이러한 변환 툴은 웹 콘텐츠를 보다 정확히 변환하기 위해 웹 서버 운영자가 수작업을 통해 변환하는 기법으로, 자동 변환이 아니며 유선 인터넷 상의 방대한 문서의 양에 비하여 변환 서비스가 제공되는 문서의 범위가 제한되는 단점이 있다.

<22> 또한 단말에서 동작하는 변환기법으로는, 줌인/줌아웃 기능을 제공하는 Smart View, Pad++ 등이 있다. 이들은 단말의 성능을 보다 정확히 파악하고 사용자 요구사항을 쉽게 반영할 수 있는 장점이 있으나, 전체 페이지를 이미지로 대략적인 정보를 확인한 후에 실제 내용 파악을 위해서는 각 부분의 줌인 인터페이스를 사용하여 다시 확대된 내용을 한번 더 확인해야 하는 불편한 점이 있다.

<23> 그리고 프락시 서버에서 동작하는 변환기법으로, 팜파일럿 단말의 브라우저를 위한 변환 프락시를 제시하는 Top Gun Wingman과 핸드헬드, 셀룰러 계열의 단말을 모두 지원하는 다이제스터(Digester) 등이 있다. 다이제스터(Digester)는 사람에 의한 직접적인 변환 수행을 통해

언은 다양한 휴리스틱 변환 기법과 이들의 적절한 적용 규칙에 따라 변환을 시도한다. 정확한 변환을 위하여 다수의 복잡한 알고리즘이 사용되고, 변환 결과 정보는 요약, 축소 또는 페이지 나눔 등으로 표현된다. 하지만 제한된 정보 표현 방법과 복잡한 카테고리 구조, 다수의 하이퍼링크 인덱스 사용으로 정보 검색에 불편한 인터페이스를 가지는 문제점이 있다.

<24> 그 외의 종래기술로는 한국 특허공개번호 2002-31691 호로 공개된 "실시간 인터넷 콘텐츠 변환 방법 및 시스템(출원번호:10-2000-0062342)"과 특허공개번호 2002-15223호로 공개된 "컨텐츠 가공 시스템 및 그 방법(출원번호:10-2000-0048415)"이 있다. 여기서 "실시간 인터넷 콘텐츠 변환 방법 및 시스템"은 미리 정해진 규칙을 이용하여 문서 내용 중 일부를 추출, 페이지 분할 또는 다른 마크업 언어로의 변환을 수행하는 것으로, 문서 축약으로 변환을 수행할 뿐 문서 분석 기법 및 재표현 방법에 대한 구체적인 내용이 없다. 또한 "컨텐츠 가공 시스템 및 그 방법"은 유선 웹 콘텐츠의 소형 단말 사용자를 위한 변환 시스템의 전반적인 구성에 대해서만 언급하고 있을 뿐이다.

<25> 따라서, 종래의 웹 문서 변환 기술은 급속한 단말 성능의 향상을 반영하지 못하고, 특정 부분만을 추출하거나 내용을 요약하는 방식의 변환과 이를 표현하기 위한 복잡한 카테고리 구조와 페이지 나눔 및 링크 연결이 주를 이루고 있고, 명확한 분석 및 변환과 표현 방법에 대한 구체적인 제안을 찾아볼 수 없다. 즉, 대부분의 선행 연구에서는 저성능의 Cellular Phone 계열의 단말을 위하여 단순 텍스트 요약 수준의 변환을 수행하였고, 근래에 고성능의 핸드헬드 단말기들이 등장했지만 여전히 변환은 내용 축약, 이미지 삭제 등의 콘텐츠 줄이기가 주가 되고 있다. 또는 페이지 분할과 링크를 이용한 페이지 연결 기법이 제공되기도 하는데, 비록 실

질적인 내용 축약은 아니지만 링크의 깊이(depth)가 깊어질 경우, 전체 내용 파악이 어려워지고 다시 이전의 페이지로 되돌아가야 하는 불편점이 있다.

【발명이 이루고자 하는 기술적 과제】

- 26> 본 발명은 이러한 문제를 해결하기 위한 것으로, 사용자의 향상된 단말의 성능을 고려하여 현재의 복잡하고 많은 정보를 내포하는 웹 문서를 원 문서의 내용을 최대한 반영하면서 동시에 편리한 인터페이스를 갖도록 변환할 수 있는 소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템 및 방법을 제공하는데 그 목적이 있다.
- 27> 상기 목적을 달성하기 위하여 본 발명의 소형 화면의 단말기를 위한 웹 문서 변환 시스템은 대형 디스플레이 화면용 웹 문서를 소형 디스플레이용 화면에 적합한 웹 문서로 변환하기 위한 웹 콘텐츠 변환시스템에 있어서, 태그 오류를 포함하는 비정형적인 웹 문서를 표준에 맞게 정제한 후 분석에 적합한 데이터 형식으로 출력하는 전처리기; 클라이언트 프로파일 정보를 추출하고 관리하는 클라이언트 프로파일 분석기; 상기 전처리기에서 정제된 웹 문서를 입력받아 문서 분석 알고리즘에 따라 웹 문서를 내용 조각 단위(컴포넌트)로 설정하는 구조분석기; 상기 웹 문

서에 포함된 이미지의 인코딩, 디코딩 과정과 이미지 크기에 대한 정보를 추출하는 이미지 변환기; 상기 정의된 내용 단위 조각(컴포넌트)을 클라이언트 성능 정보와 내용 단위 조각(컴포넌트)의 속성 값을 이용하여 단말 화면의 최대폭을 초과하지 않는 범위 내에서 유사한 조각들로 그룹화하는 컴포넌트 블록 추출기; 상기 컴포넌트 블록 추출기에 의해 생성된 각 컴포넌트 블록에 대하여 포함되는 콘텐츠의 특성에 따라서 인덱스와 본문 내용 부분으로 분류하는 컴포넌트 블록 카테고리부; 상기 인덱스로 분류된 컴포넌트 블록으로부터 이미지 또는 텍스트 인덱스 정보를 추출하고, 이를 표현하기 위한 스크립트 파일 및 추가 태그 집합을 생성하는 인덱스 생성기; 음성 지원 기능을 수행하기 위해 텍스트 위주의 본문 내용 블록에 대하여 음성 마크업 언어로 변환하는 음성 마크업 생성기; 및 상기 생성된 내용 객체 요소들을 문서 양식에 따라 적절히 재배치하고 재구성하여 소형 디스플레이 화면에 적합한 웹 문서를 생성하는 커스텀화된 HTML 생성기;를 포함하는 것을 특징으로 한다.

- <28> 또한 상기 목적을 달성하기 위하여 본 발명의 변환 방법은, 대형 디스플레이 화면용 웹 문서를 소형 디스플레이용 화면에 적합한 웹 문서로 변환하기 위한 웹 콘텐츠 변환 방법에 있어서, 태그 오류를 포함하는 비정형적인 웹 문서를 표준에 맞게 정제한 후 분석에 적합한 데이터 형식으로 출력하는 전처리 단계; 상기 전처리 단계에서 정제된 웹 문서를 입력받아 문서 분석 알고리즘에 따라 태그를 분석하여 웹 문서를 내용 조각 단위(컴포넌트)로 설정하는 웹 문서 분석단계; 상기 정의된 내용 단위 조각(컴포넌트)을 클라이언트 성능 정보와 내용 단위 조각(컴포넌트)

의 속성 값을 이용하여 최대폭을 초과하지 않는 범위 내에서 유사한 조각들로 그룹화하는 컴포넌트 블록 설정 단계; 상기 컴포넌트 블록 추출 단계에 의해 생성된 각 컴포넌트 블록에 대하여 포함되는 콘텐츠의 특성에 따라서 인덱스와 본문 내용 부분으로 분류하는 컴포넌트 블록 분류 단계; 상기 인덱스로 분류된 컴포넌트 블록으로부터 이미지 또는 텍스트 인덱스 정보를 추출하고, 이를 표현하기 위한 스크립트 파일 및 추가 태그 집합을 생성하는 인덱스 생성 단계; 음성 지원 기능을 수행하기 위해 텍스트 위주의 본문 내용 블록에 대하여 음성 마크업 언어로 변환하는 음성 마크업 생성 단계; 및 상기 생성된 내용 객체 요소들을 문서 양식에 따라 적절히 재배치하고 재구성하여 소형 디스플레이 화면에 적합한 웹 문서를 생성하는 HTML 생성 단계;를 포함하는 것을 특징으로 한다.

<29> 이러한 구성 및 방법에 의하면, 본 발명은 기존의 정보 추출 및 요약의 방식이 아닌 내용 단위 블록별 재배치를 통해서 현재의 복잡하고 많은 정보를 한꺼번에 표현하는 웹 문서의 특성을 반영하고, 다수의 깊이(depth)를 가지는 인덱스 구조 또는 페이지 나눔의 방식이 아니라 내용 단위 블록의 분류와 인덱스 생성, 그리고 음성지원 문서 형식으로서의 변환을 통하여 좌우 스크롤이 없이 시청각적 표현을 동시에 지원하는 편리한 인터페이스를 제공하는 특징을 가진다.

<30> 따라서 본 발명은 단말의 화면 크기를 고려한 내용 단위 블록의 재배치, 인덱스 블록의 추출 및 다양한 인덱스 생성 기능을 통하여 전체 웹 문서를 좌우 스크롤없이 브라우징할 수 있게 하고, 텍스트 위주의 내용 본문 블록의 경우는 음성이 지원되는 마크업 언어로 변환되어 보다 편리한 인터페이스를 제공할 수 있고, 전체 구조를 소형 화면에 맞게 구성함으로써 원본 웹 문서의 내용을 최대한 반영할 수 있다.

【발명의 구성 및 작용】

- <31> 이하, 본 발명의 바람직한 실시예를 첨부된 도면을 참조하여 설명한다. 도 1은 시각적인 분리 및 그룹화로 상이한 내용 블록을 표현하는 웹 문서의 예시도이다.
- <32> 도 1을 참조하면, 웹 문서는 HTML을 작성하는 제작자가 내용을 명확하게 전달하기 위해서 의미상의 차이를 가지는 내용에 대해서 레이아웃 및 구조적 태그를 사용하여 시각적인 분리가 이루어지도록 디자인한다. 이러한 시각적인 분리는 "TABLE" 등의 구조적 표현을 위한 태그를 이용하는 경우가 많고, 따라서 이러한 태그들을 분석함으로써 전체 구조를 파악할 수 있다. 이때 일부 무분별한 태그 집합의 사용과 HTML 자체가 가지는 구조와 의미의 불명확한 구분을 고려하여, 구조적 태그뿐 아니라 태그의 속성값, 태그가 가지는 데이터의 특성, 태그 객체의 데이터 정보가 표현되는 위치정보 등을 활용하여 분석한다.
- <33> 이러한 웹 문서의 구조 분석을 통하여, 도 1과 같은 시각적 분리 레이아웃을 구성하는 최소 내용 단위 조각(이를 '컴포넌트(Component)'라 한다)(101)을 설정하고, 내용 단위 조각(101)을 사용자 단말의 성능, 특히 디스플레이 성능을 고려하여 그룹화하고 이를 내용 단위 블록(이를 '컴포넌트 블록(Component Block)'이라 한다)(102)으로 표현한다.
- <34> 이 내용 단위 블록(102)들은 포함하는 콘텐츠의 특성에 따라서 '인덱스' 부분과 '내용 본문' 부분으로 분류되고, 각각 적절한 형식으로 재표현되어진다. 인덱스 부분의 경우 후술하는 도 9a의 121과 같이 상단의 선택 박스 형식으로 재표현 되어지고, 본문 부분으로 분류된 경우는 도 9a의 122와 같이 주요 내용 부분으로 변환없이 재배치만 되거나 도 9b의 123과 같이 음성이 지원되는 문서 형식으로 변환되어 표현된다.

- <35> 도 2는 본 발명에 따른 소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템의 모듈 구성 개념도이고, 도 3은 본 발명에 따른 소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템의 동작 과정 순서도이다.
- <36> 본 발명에 따른 콘텐츠 변환 시스템의 구성은 도 2에 도시된 바와 같이, 전처리 단계(S1), 웹 문서 분석단계(S2), 웹 문서 변환단계(S3), 웹 문서 생성단계(S4)로 분리되고, 각 단계(S1~S4)는 그 동작을 처리하는 세부 모듈들(201~209)로 이루어진다.
- <37> 전처리 단계(S1)는 전처리기(PreProcessor: 201)와, 클라이언트 프로파일 분석기(Client Profile Analyzer: 202)로 이루어진다. 전처리기(PreProcessor: 201)는 태그 오류를 포함하는 비정형적인 웹 문서를 표준에 맞게 정제한 후 분석에 적합한 데이터 형식으로 출력한다. 클라이언트 프로파일 분석기(Client Profile Analyzer: 202)는 클라이언트 정보를 받아들이는 기능을 수행한다. 클라이언트 정보는 HTTP Header 필드에 포함시켜 전송하거나 또는 특정 통신 규약을 활용하여 정보를 전송할 수 있다. 그 외에 외부 모듈과의 입출력 관리 부분을 전처리단계(S1)에서 수행한다.
- <38> 웹 문서 분석단계(S2)의 구조분석기(Layout-Based Structure Analyzer: 203)는 전처리 단계(S1)에서 정제된 웹 문서를 입력받고, 문서 분석 알고리즘을 통하여 웹 문서를 내용 조각 단위(Component)로 설정한다. 이미지 변환기(Image Converter: 204)는 웹 문서의 이미지 인코딩, 디코딩 과정과 이미지 크기에 대한 정보를 추출한다.
- <39> 웹 문서 변환단계(S3)의 컴포넌트 블록 추출기(Component Block Extractor: 205)는 정의된 내용 단위 조각(컴포넌트)을 클라이언트 성능 정보와 내용 단위 조각(컴포넌트)의 속성 값을 이용하여 단말 화면의 최대폭(MAX_WIDTH)을 초과하지 않는 범위 내에서 유사한 조각들로 그룹화한다. 컴포넌트 블록 카테고리부(Component Block Categorizer: 206)는 각 컴포넌트 블록

(Component Block)에 대하여 포함하는 컨텐츠의 특성에 따라서 '인덱스'와 '본문 내용' 부분으로 분류한다.

40> 웹 문서 생성단계(S4)에서는 필요한 내용 객체들을 생성하는 과정을 수행한다. 인덱스 생성기(Index Generator:207)는 인덱스로 분류된 컴포넌트 블록(Component Block)으로부터 이미지(Image) 또는 텍스트(Text) 인덱스 정보를 추출하고, 이를 표현하기 위한 스크립트 파일 및 추가 태그 집합을 생성한다. 음성 마크업 생성기(Auditory Markup Generator: 208)는 음성 지원 기능을 수행하기 위해 텍스트 위주의 본문 내용 블록에 대하여 VoiceXML 등의 마크업 언어로 변환하는 과정을 수행한다. 이때 브라우저에서는 음성 정보의 웹 문서를 소리로 랜더링하는 기능을 제공해야 한다. 마지막으로 커스텀화된 HTML 생성기(Customized HTML Generator: 209)는 이전 단계를 통하여 생성된 내용 객체 요소들을 문서 양식에 따라 적절히 재배치하고 재구성하여 커스텀화된 웹 문서를 생성한다.

41> 도 3은 도 2의 전체 동작 과정을 설명하기 위한 순서도이다. 도면을 참조하면, 원본 HTML 파일을 입력 받아서 HTML 문서를 정제하는 단계를 거쳐서 HTML DOM 트리(Tree) 형식의 데이터 구조를 출력한다(401~403). 이는 도 2의 전처리기(201) 모듈에서 동작한다. 웹 문서 분석(HTML 태그 분석) 단계(404)에서는 트리(Tree) 데이터를 입력으로 받아서 태그 분석을 수행하고, 이 과정은 도 2의 구조분석기(203)와 이미지 변환기(204)에서 동작한다. 웹 문서 분석 단계(404)의 세부 알고리즘은 도 4의 순서도를 참고하여 아래에서 설명한다.

42> 태그 분석 후 다음 단계인 컴포넌트 블록(Component Block) 설정 단계(405)는 도 2의 컴포넌트 블록 추출기(205)에서 동작하고, 그 다음의 컴포넌트 블록(Component Block) 분류(406) 단계는 도 2의 컴포넌트 블록 카테고리부(206)에서 동작한다. 컴포넌트 블록(Component

Block) 설정(405)과 컴포넌트 블록(Component Block) 분류(406) 단계는 도 6과 도 7의 순서도를 참고하여 각각의 알고리즘을 설명한다.

<43> 먼저, 도 5를 참조하여 웹 문서 분석 단계(404)의 세부 알고리즘을 설명하면 다음과 같다.

<44> 본 발명의 분석 알고리즘에서는 <TABLE>, <TR>, <TD>, 등의 태그를 주로 사용하고, 특정 <TD> 태그를 컴포넌트(Component)로 정의하여 내용 단위 분석의 최소 단위로 사용한 경우에 대해 설명한다.

<45> 먼저 HTML Document Tree 데이터 구조를 입력으로 받고, 사용자 단말이 받아들이는 최대 화면폭(width)을 최대폭 "MAX_WIDTH"로 정의한다(501,502). 분석 과정을 거치면서 <TD> 태그 노드에는 아래의 표 1과 같은 정보가 추가로 저장되고, 이후 컴포넌트 블록(Component Block) 추출에 이용된다.

<46> 【표 1】

변수	내용
width	픽셀 단위로 재계산된 width 값
Comp_num	component 로 설정되는 경우 component의 ID를 표현하는 값 일반 component : (sequence number,0,0) nested component : (0,첫째자식의 Comp_num의 첫번째 숫자, 마지막 자식의 Comp_num의 첫번째 숫자)
Col_num	전체 테이블 구조의 레이아웃에서 몇 번째 Column에 위치하는가를 나타내는 숫자
Row_num	전체 테이블 구조의 레이아웃에서 몇 번째 Row에 위치하는가를 나타내는 숫자
Table_depth	<TD>가 가지는 조상 <table> 태그 노드의 개수 즉, nested_table의 depth를 나타냄.

<47> 단계 502에서 전역변수에 대한 초기화가 끝난 후, 모든 태그 노드(Tag Nodes)를 프리오더(PreOrder) 순으로 방문하면서 아래의 과정을 반복한다(503).

- <48> 만약 방문한 노드가 <TABLE>인 경우(504)에는 테이블 깊이(Table_Depth)를 확인(505)하고 경계값(예컨대, 3)을 초과할 경우 <Table> 태그와 그 아래의 모든 자손 노드를 일반 콘텐츠로 간주하여 폭(width) 설정 과정(506)만 거치고 더 이상의 분석은 수행하지 않는다. 테이블 깊이(Table_depth)가 경계값(예컨대, 3)을 초과하지 않을 경우는 테이블 깊이(Table_depth)값을 하나 증가(507)시킨다.
- <49> 만약, 방문한 노드가 <TR>인 경우(508)에는 행번호(Row_num)를 증가시킨다(509). 단, 네스티드 테이블(nested table)의 첫번째 행(row)일 경우는 증가시키지 않는다. 또한 루트 테이블(root table)의 <TR> 태그일 경우에는 열번호(Col_num)를 0으로 초기화시킨다.
- <50> 만약, 방문한 노드가 <TD>인 경우(510)에는 콘텐츠를 가지는가를 판단(511)하여 열번호(Col_num)를 증가(512)시킨다. 단, 네스티드 테이블(nested table) <TR>의 첫번째 <TD>는 증가시키지 않는다. 만약 <TD>가 내용 콘텐츠를 포함하지 않고 레이아웃 표현을 위해서 사용되었을 경우, 폭(width) 설정 단계(522)로 넘어가고 내용 콘텐츠를 포함하는 경우에는 컴포넌트(Component)로 설정하고 구조적 정보를 추가한다.
- <51> 즉, 컴포넌트(Component)는 콘텐츠를 가지는 <TD> 태그 블록으로 정의된다. 이러한 컴포넌트(Component) 중에서 자손으로 <TABLE> 태그를 가지면(513) 네스티드 컴포넌트(nested component)로 설정하여 컴포넌트 번호(Comp_num)값을 상기 표 1에서와 같이 표기(514)하고, <TABLE> 이외의 다른 태그들을 콘텐츠로 가질 경우에는 일반 컴포넌트(component)로 설정하여 증가된 순번(sequence number)으로 컴포넌트 번호(Comp_num) 변수를 정의(515)한다.

- <52> 도 3의 TABLE 태그의 표현 계층 관계도를 참조하면 <TD> 태그가 포함할 수 있는 태그 종류를 확인할 수 있다. 도면을 참조하면, 테이블(TABLE)은 TR과 CAPTION으로 나뉘고, TR은 TH와 TD로 구분된다.
- <53> 방문한 노드가 인 경우(516)에는 폭(width)을 확인한 후 변경한다(517,518). 만약, 폭(width)이 변경된다면 이미지 맵이 설정되었는지 확인하고, 설정되었다면 좌표값을 나타내는 이미지 맵 코드 <AREA>의 COORDS 속성값을 520의 수식을 이용하여 수정한다. 단계 518의 폭(width) 설정 과정은 %로 설정된 값을 픽셀로 환산하고, 최대폭(MAX_WIDTH)을 초과할 경우 최대폭(MAX_WIDTH)으로 대체하며, 만약 폭(width) 속성 값이 설정되어 있지 않다면 <TR> 폭(width), <TD> 폭(width)의 합, 최대 폭(width)등을 이용하여 유추한다.
- <54> 도 7a와 도 7b의 예제를 통하여 도 5의 알고리즘으로부터 얻은 구조적 정보를 확인해본다.
- <55> 도 7a는 구조적 태그의 시각적인 모습을 표현한 것으로, <TABLE>, <TR>, <TD> 블록을 표시하고 내용을 가지는 <TD> 태그 블록이 대해 컴포넌트(Component)를 설정한다. 추가되는 정보는 아래의 표 2에서 보여준다. 도 7b는 도 7a와 같은 태그 집합에 대해 구조적 태그의 트리 모형으로 표현한 것으로, 태그간의 계층 관계를 쉽게 파악할 수 있다.
- <56>

【표 2】

(A)	Comp_num	Row_num	Col_num	Table_depth	Width
①	(1,0,0)	1	1	1	200
②	(2,0,0)	1	2	1	400
③	(3,0,0)	1	3	1	200
④	(0,4,7)	2-5	1-1	1	150
⑤	(4,0,0)	2	1	2	150
⑥	(5,0,0)	3	1	2	150
⑦	(6,0,0)	4	1	2	150
⑧	(7,0,0)	5	1	2	150
⑨	(0,8,15)	2-5	2-4	1	650->MAX_WIDTH
⑩	(8,0,0)	2	2	2	650->MAX_WIDTH
⑪	(0,9,14)	3-5	2-3	2	400
⑫	(9,0,0)	3	2	3	200
⑬	(10,0,0)	3	3	3	200
⑭	(11,0,0)	4	2	3	200
⑮	(12,0,0)	4	3	3	200
⑯	(13,0,0)	5	2	3	200
⑰	(14,0,0)	5	3	3	200
⑱	(15,0,0)	3	4	3	250
16	(16,0,0)	6	1	1	800->MAX_WIDTH

57> 상기 표 2에서 (A)는 도 7a,b에 표기된 컴포넌트 번호(Comp_num)의 첫번째 숫자이고, 최대폭(MAX_WIDTH)은 500 픽셀 미만으로 가정한다.

58> 다음으로, 컴포넌트 블록(Component Block)은 컴포넌트(Component) 단위를 기준으로 그 내부에 포함되는 모든 태그 집합을 별도의 <TABLE> 태그의 단일<TD>로 묶어서 상위 조상 <TABLE>과 동등한 위치에 삽입하여 생성한다.

59> 도 6과 도 7b를 참고로 컴포넌트 블록 설정(405) 과정의 세부 알고리즘을 설명하면 다음과 같다.

60> 먼저 컴포넌트 트리(Component_tree)를 입력받아 모든 컴포넌트 노드(Component Node)에 대해 초기 폭(width) 정보를 확인하고 최대폭(MAX_WIDTH)을 초과할 때 그 다음 과정을 수행한다(601-603). 현재 컴포넌트(component) 노드(A)의 형제 노드가 있는지 확인하고, 있다면 최대

폭(MAX_WIDTH)이 넘지 않는 범위 내에서 유사한 형제 노드들을 묶는 그룹화 과정을 수행한다 (605~607). 도 7b의 예에서 ①, ②, ③의 컴포넌트(Component)는 (①), (②), (③) 또는 (①③), (②)의 그룹으로 만들 수 있다.

- <61> 다음의 테이블 블록화 단계(608)에서 각 그룹에 속하는 모든 태그 집합을 '<TABLE><TR> Component①, ③</TR></TABLE>'과 같은 형식으로 하나의 테이블 블록으로 표현한다. 또는 형제 노드가 없다면 단계 608의 컴포넌트(Component) 노드의 테이블 블록화 과정만 거친다.
- <62> 단계 609의 테이블 블록의 재배치 단계에서는 상위 과정에서 새로이 생성된 테이블 블록을 (A)의 부모의 부모 노드인 <TABLE> 노드(B)의 이전 형제 노드(previous sibling node)로 삽입한다.
- <63> 만약, (A)가 (B)의 마지막 <TD>노드이고, (B)가 네스티드 테이블(nested table) 이면 다음 단계로 진행하고, 아니면 단계 602에서 다음 노드를 방문하여 앞의 과정을 반복한다.
- <64> 다음 단계로 진행하는 경우는 도 7b의 ⑦, ⑭, ⑮가 (A), 즉 현재 방문중인 컴포넌트(Component)일 때가 된다. (B)를 자손으로 가지는 상위 조상 <TD> 즉, (C)가 네스티드 컴포넌트(nested Component)일 경우 단계 609의 과정을 수행한다. 즉, 도 7b의 ⑦, ⑭가 이에 해당되고 각각의 (C)는 ㉠과 ㉡가 된다. (C)의 자식 노드 중에서 (B)를 포함하는 자식노드(도 7b의 701)를 기준으로 좌측과 우측의 모든 형제 노드를 각각 테이블 블록(도 7 b의 702, 703)으로 묶는다. 다시 (C)를 포함하는 테이블 블록을 만들고(614), 단계 609의 과정을 반복한다.
- <65> 컴포넌트(Component)는 테이블 블록화를 통하여 하나의 표현 단위로 추출되고 이를 컴포넌트 블록(Component Block)으로 정의한다. 각각의 컴포넌트 블록(Component Block)은 트리상

에서의 컴포넌트(Component)의 위치에 따라 배치순서가 정해지고, 순서에 따라 위에서 아래로 테이블 블록의 모양으로 표현되어진다.

<66> 계속해서 도 8을 참고로 컴포넌트 블록(Component Block)의 분류 과정(406)의 세부 알고리즘을 설명한다.

<67> 컴포넌트 블록 트리를 입력받아 모든 컴포넌트 블록(Component Block)을 방문하면서 컴포넌트 블록(Component Block)의 콘텐츠 패턴을 비교한다(801~803). 이때 사용 가능한 비교 변수로는 아래의 표 3에서 정리한다.

<68> 【표 3】

변수	기대되는 패턴
Text_Length	유사 반복, 제한된 짧은 길이
Image_Width	유사 반복, 제한된 width
Link_Number	거의 모든 콘텐츠가 링크정보를 가짐. 연결된 문서의 위치, 파일명의 유사성 비교
Row_num	작은 수로 제한. 웹문서에서 상단에 배치된 블록으로 제한
Col_num	최대 또는 최소값으로 제한, 좌측 또는 우측에 배치된 블록으로 제한

<69> 패턴 비교 결과값이 임의의 경계값을 초과하는지 그 여부에 따라 초과하면 인덱스 타입(INDEX type)으로 결정한다(804,805). 인덱스(INDEX)로 결정된 컴포넌트 블록은 그 콘텐츠의 데이터 타입이 이미지인지 또는 텍스트인지에 따라서 각각 이미지 인덱스(INDEX_I), 텍스트 인덱스(INDEX_T)로 타입(type) 값을 설정한다(806~808).

<70> 인덱스(INDEX)가 아닌 블록은 본문(BODY)으로 구분되고, 포함하는 콘텐츠에서 텍스트의 비중에 따라, 음성(Voice) 지원 문서로의 변환을 수행하게 되는 음성 본문(BODY_V) 타입과 그

외의 일반 내용 블록으로 처리되는 일반 본문(BODY_G)으로 구분된다(809~812). 단계 813에서 마지막 블록이 아니면 다음 블록에 대해서 단계 802부터 반복한다.

- <71> 상기 분류 이후 과정은 도 4의 전체 동작 과정을 보인 순서도를 참고하여 설명한다.
- <72> 도면을 참조하면, 컴포넌트 블록(Component Block)이 분류된 후 각 컴포넌트 블록의 타입(Type)에 따라 도 4의 단계 411, 413, 414를 거치거나 또는 컴포넌트 블록을 그대로 추출한다. 이 과정을 모든 컴포넌트 블록(Component Block)에 대해 수행하고, 마지막의 단계 416에서 각 블록들을 적절히 배치하여 새로운 HTML 문서(417)를 생성한다. 컴포넌트 블록(Component Block)의 타입(type) 별 동작과정을 살펴보면, 다음과 같다.
- <73> 만일, 컴포넌트 블록의 타입이 음성 본문(BODY_V)(Type = BODY_V)이라면, 단계 411의 음성(Voice) 문서 생성 단계를 거쳐 음성 지원 문서를 생성한다. 이는 도 2의 음성 마크업 생성기(208) 모듈에서 동작하며, 블록(Block)에서 모든 텍스트 부분은 다음 표 4의 샘플 코드에서와 같이 <prompt> value로 추가하여 간단한 VoiceXML 문서를 생성할 수 있다. 생성된 문서는 별도의 파일로 저장되며 원본 HTML에서 링크로 연결된다.

<74> 【표 4】

```
<?xml version="1.0"?>
<vxml version="1.0">
<form>
<block>
<prompt>
```

(BODY_V로 분류된 Block에서 추출한 text 정보를 value로 추가한다)

- <75> 여기서, 컴포넌트 블록의 타입이 일반 본문(BODY_G)(Type = BODY_G)이라면, 일반적인 콘텐츠 요소이므로 그대로 추출하여 재배치한다.

<76> 만일, 컴포넌트 블록의 타입이 이미지 인덱스(INDEX_I)(Type == INDEX_I)라면, 이미지 인덱스 생성 과정(413)을 통하여 자바 스크립트(Java Script)로 표현되는 이미지 인덱스(Image Index)를 생성한다. 다음 표 5의 샘플 코드 예제와 같이 간단한 스크립트(script) 파일을 자동 생성하고, 이미지 파일을 매핑하여 구현 할 수 있다.

<77> 【표 5】

```
// HEAD에 들어갈 자바 스크립트
<SCRIPT LANGUAGE="JavaScript">
<!--
image1= new Image();
image1.src = "image1.gif";
image2= new Image();
image2.src = "image2.gif";
image3= new Image();
image3.src = "image3.gif";
image4= new Image();
image4.src = "image4.gif";
links = new Array;
links[0] = "LINK #1";
links[1] = "LINK #2";
links[2] = "LINK #3";
links[3] = "LINK #4";
function imgchange(){
    var imageNum = document.form.selImage.selectedIndex + 1;
    fname = eval("image" + imageNum + ".src");
    document.img.src = fname;
}
```

<78> 여기서, 컴포넌트 블록의 타입이 텍스트 인덱스(INDEX_T)(Type == INDEX_T)라면, 인덱스 정보가 텍스트로 표현되어지는 것으로, 텍스트 인덱스 생성 과정(414)을 통하여 다음 표 6과 같이 <select> 태그를 사용하여 재표현한다. 이미지 인덱스 생성(413) 과정과 텍스트 인덱스 생성(414) 과정은 모두 도 2의 인덱스 생성기(207) 모듈에서 동작하고, 인덱스 정보의 추출은 통상의 방식으로 구현할 수 있다.

<79> 【표 6】

```
// HEAD 에 들어갈 자바스크립트
<script language="JavaScript">
<!--
    function change(form){
        var list=form.selectedIndex;
        location type=form.options[list].value;

// location type을 다음의 것들 중에서 선택
// - self.location.href : 자기 자신이 속해있는 프레임으로 링크
// - top.location.href : 프레임과 무관하게 전체 화면이 모두 바뀜
// - parent.location.href : 자기 자신을 포함하고 있는 부모 프레임이 바뀜
// - parent.frameName.location.href : 부모프레임 중 선택된 이름을 가진 자식 프레임으로 링크
    form.selectedIndex = 0;
```

<80> 위와 같이 각 컴포넌트 블록을 내용 특성에 따라서 적절한 방법으로 표현한 후, 도 2의 HTML 생성기(209)에서 동작하는 새로운 HTML 구성 및 생성 단계(416)에서 이들 내용 객체를 배치하고 생성하는 과정을 거친다. 다음 표 7의 샘플 코드는 전체 HTML의 태그 구성과 각 내용 객체들의 간단한 배치 방법을 제시한다.

<81> 【표 7】

```
<HTML>
<HEAD>
<TITLE></TITLE>
<SCRIPT> --> Java Script Generator 모듈에 의해 자동 생성된 스크립트 파일을 첨부한다. 이것은 Image Index가 생성되는 경우 추가된다.
</SCRIPT>
</HEAD>
<BODY> --> INDEX_T 또는 BODY_G로 분류되어진 Component Block을 BODY 태그 안에 붙인다.
<SELECT>
<OPTION> --> Text Index의 수만큼 select list form을 생성하고 각각의 value 값을 Option태그로 적절히 배치한다.
</SELECT>
```

<82> 이상에서 설명한 바와 같은 본 발명의 콘텐츠 변환 시스템은 웹 서버, 클라이언트, 그리고 프락시의 세가지 레이어(Layer)에 놓여질 수 있고, 각각은 환경에 따른 장단점을 가진다.

또한 컴포넌트 및 컴포넌트 블록의 추출 알고리즘의 구현은 다양한 방법으로 이루어 질 수 있으며, 인덱스 생성 및 음성 문서의 생성 방법 또한 여러 가지 구현 방법 중의 한가지를 예제로 보이고 있다.

83> 도 9a, 9b는 본 발명에 따른 웹 콘텐츠 변환 결과를 도시한 예시도이다.

84> 도 9a는 내용 단위 객체의 재배치와 인덱스 추출을 통해 변환된 웹 문서의 결과 페이지이고, 도 9b는 여기에 음성지원 마크업 생성 기능이 추가되었을 경우 나타나는 결과 페이지를 표현한다.

【발명의 효과】

85> 상술한 바와 같이, 본 발명은 소형 화면의 단말을 가진 사용자가 무선 인터넷에 접속하여 웹 서비스를 사용하고자 할 경우, 기존의 일반 데스크탑 PC의 디스플레이 성능에 적합하도록 작성된 웹 문서를 소형 화면에서도 효율적으로 표현되어질 수 있도록 변환해 주기 위한 새로운 기법 및 시스템을 제시한다. 본 발명에 따라 웹 문서는 구조적 태그 정보의 분석으로 내용 단위 조각이 설정되고, 유사 내용 단위 그룹으로 묶여진 후, 콘텐츠 내용 정보를 기반으로 인덱스 또는 본문 내용으로 분류되어서 각각을 재배치하여 전체 웹 페이지에 대해 좌우 스크롤 없는 편리한 인터페이스로 브라우징할 수 있는 기능을 제공한다. 또한 인덱스의 추출 및 생성 그리고 음성지원 웹 문서로의 변환도 함께 제공하여 웹 문서의 다양한 재구성 및 소형 단말의 특성을 고려한 표현 효과를 가진다. 또한 원 문서의 내용을 최대한 유지하여 의미 전달을 명확하게 하는 효과도 얻을 수 있다.

- 36> 이상에서 설명한 것은 본 발명에 따른 소형 화면 단말기를 위한 웹 콘텐츠 변환 방법 및 시스템을 설명한 하나의 실시 예에 불과한 것으로써, 본 발명은 상기한 실시 예에 한정되지 않고, 이하의 특허 청구의 범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변경 실시가 가능한 범위까지 본 발명의 기술적 사상이 미친다고 할 것이다.

【특허청구범위】

【청구항 1】

대형 디스플레이 화면용 웹 문서를 소형 디스플레이용 화면에 적합한 웹 문서로 변환하기 위한 웹 콘텐츠 변환시스템에 있어서,

태그 오류를 포함하는 비정형적인 웹 문서를 표준에 맞게 정제한 후 분석에 적합한 데이터 형식으로 출력하는 전처리기;

클라이언트 정보를 추출하고 관리하는 클라이언트 프로파일 분석기;

상기 전처리기에서 정제된 웹 문서를 입력받아 문서 분석 알고리즘에 따라 웹 문서의 내용 조각 단위(컴포넌트)로 설정하는 구조분석기;

상기 웹 문서에 포함된 이미지의 인코딩, 디코딩 과정과 이미지 크기에 대한 정보를 추출하는 이미지 변환기;

상기 정의된 내용 단위 조각(컴포넌트)을 클라이언트 성능 정보와 내용 단위 조각(컴포넌트)의 속성 값을 이용하여 최대폭을 초과하지 않는 범위 내에서 유사한 조각들로 그룹화하는 컴포넌트 블록 추출기;

상기 컴포넌트 블록 추출기에 의해 생성된 각 컴포넌트 블록에 대하여 포함되는 콘텐츠의 특성에 따라서 인덱스와 본문 내용 부분으로 분류하는 컴포넌트 블록 카테고리부;

상기 인덱스로 분류된 컴포넌트 블록으로부터 이미지 또는 텍스트 인덱스 정보를 추출하고, 이를 표현하기 위한 스크립트 파일 및 추가 태그 집합을 생성하는 인덱스 생성기;

음성 지원 기능을 수행하기 위해 텍스트 위주의 본문 내용 블록에 대하여 음성 마크업 언어로 변환하는 음성 마크업 생성기; 및

상기 생성된 내용 객체 요소들을 문서 양식에 따라 적절히 재배치하고 재구성하여 소형 디스플레이 화면에 적합한 웹 문서를 생성하는 HTML 생성기;를 포함하는 것을 특징으로 하는 소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템.

【청구항 2】

제 1항에 있어서, 상기 콘텐츠 변환 시스템은 웹 서버, 클라이언트, 및 프락시의 세가지 레이어 중 어느 하나에 설치될 수 있는 것을 특징으로 하는 소형 화면 단말기를 위한 웹 콘텐츠 변환 시스템.

【청구항 3】

대형 디스플레이 화면용 웹 문서를 소형 디스플레이용 화면에 적합한 웹 문서로 변환하기 위한 웹 콘텐츠 변환 방법에 있어서,

태그 오류를 포함하는 비정형적인 웹 문서를 표준에 맞게 정제한 후 분석에 적합한 데이터 형식으로 출력하는 전처리 단계;

상기 전처리 단계에서 정제된 웹 문서를 입력받아 문서 분석 알고리즘에 따라 태그를 분석하여 웹 문서를 내용 조각 단위(컴포넌트)로 설정하는 웹 문서 분석단계;

상기 정의된 내용 단위 조각(컴포넌트)을 클라이언트 성능 정보와 내용 단위 조각(컴포넌트)의 속성 값을 이용하여 최대폭을 초과하지 않는 범위 내에서 유사한 조각들로 그룹화하는 컴포넌트 블록 설정 단계;

상기 컴포넌트 블록 추출 단계에 의해 생성된 각 컴포넌트 블록에 대하여 포함되는 콘텐츠의 특성에 따라서 인덱스와 본문 내용 부분으로 분류하는 컴포넌트 블록 분류 단계;

상기 인덱스로 분류된 컴포넌트 블록으로부터 이미지 또는 텍스트 인덱스 정보를 추출하고, 이를 표현하기 위한 스크립트 파일 및 추가 태그 집합을 생성하는 인덱스 생성 단계;

음성 지원 기능을 수행하기 위해 텍스트 위주의 본문 내용 블록에 대하여 음성 마크업 언어로 변환하는 음성 마크업 생성 단계; 및

상기 생성된 내용 객체 요소들을 문서 양식에 따라 적절히 재배치하고 재구성하여 소형 디스플레이 화면에 적합한 웹 문서를 생성하는 HTML 생성 단계;를 포함하는 것을 특징으로 하는 소형 화면 단말기를 위한 웹 콘텐츠 변환 방법.

【청구항 4】

제 3 항에 있어서, 상기 웹 문서 분석단계는 <TABLE>, <TR>, <TD>, 등의 태그를 주로 분석하고, 특정 <TD> 태그를 컴포넌트(Component)로 정의하여 내용 단위 분석의 최소 단위로 사용한 것을 특징으로 하는 소형 화면 단말기를 위한 웹 콘텐츠 변환 방법.

【청구항 5】

제 3항에 있어서, 상기 컴포넌트 블록 설정단계는

컴포넌트 트리(Component_tree)를 입력받아 모든 컴포넌트 노드(Component Node)에 대해 초기 폭 정보를 확인하고, 현재 컴포넌트 노드의 형제 노드가 있는지 확인하고, 있다면 최대 폭(MAX_WIDTH)을 넘지 않는 범위 내에서 유사한 형제 노드들을 묶어 그룹화하는 것을 특징으로

하는 소형 화면 단말기를 위한 웹 콘텐츠 변환 방법.

【청구항 6】

제 3항에 있어서, 상기 컴포넌트 블록 분류단계는

컴포넌트 블록 트리를 입력받아 모든 컴포넌트 블록(Component Block)을 방문하면서 컴포넌트 블록(Component Block)의 콘텐츠 패턴을 비교하는 단계;

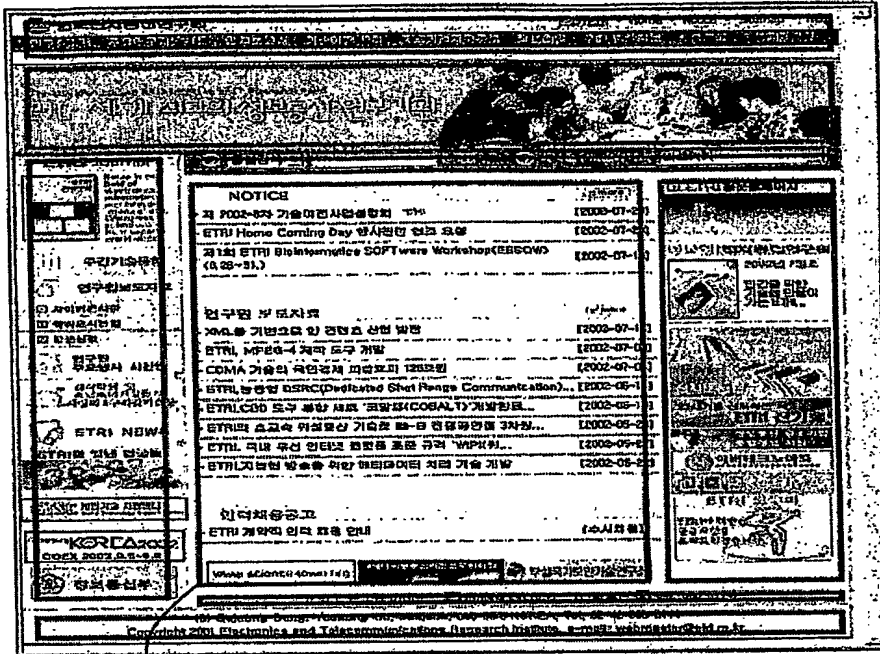
상기 패턴 비교 결과값이 임의의 경계값을 초과하는지 그 여부에 따라 초과하면 인덱스 타입을 결정하는 단계;

인덱스로 결정된 블록은 그 콘텐츠의 데이터 타입이 이미지인지 또는 텍스트인지에 따라서 각각 이미지 인덱스(INDEX_I)나 텍스트 인덱스(INDEX_T)로 타입을 설정하는 단계; 및

인덱스(INDEX)가 아닌 블록은 본문(BODY)으로 구분하고, 포함되는 콘텐츠에서 텍스트의 비중에 따라 음성(Voice) 지원 문서로의 변환을 수행하게 되는 음성본문(BODY_V) 타입과 그 외의 일반 내용 블록으로 처리되는 일반본문(BODY_G)으로 구분하는 단계;로 이루어지는 것을 특징으로 하는 소형 화면 단말기를 위한 웹 콘텐츠 변환 방법.

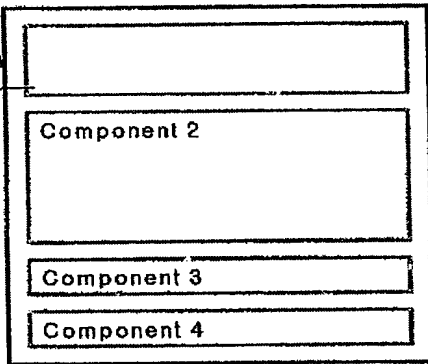
【도면】

【도 1】

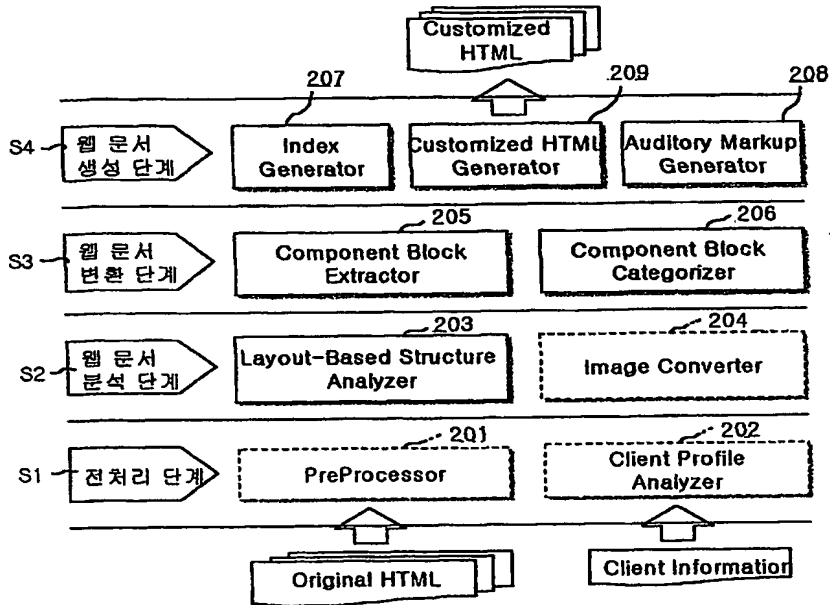


101

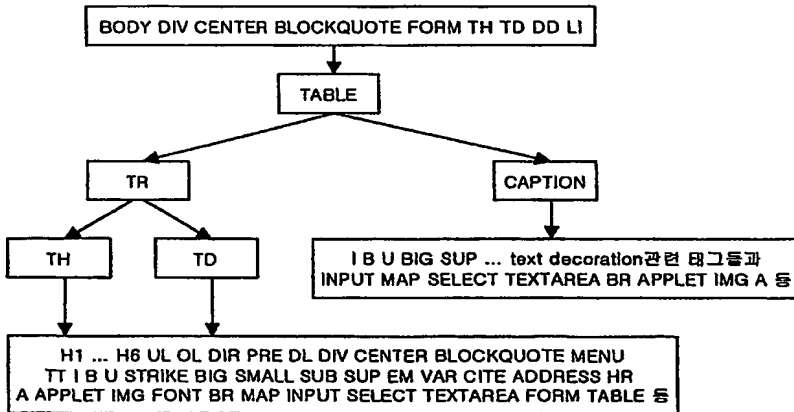
102



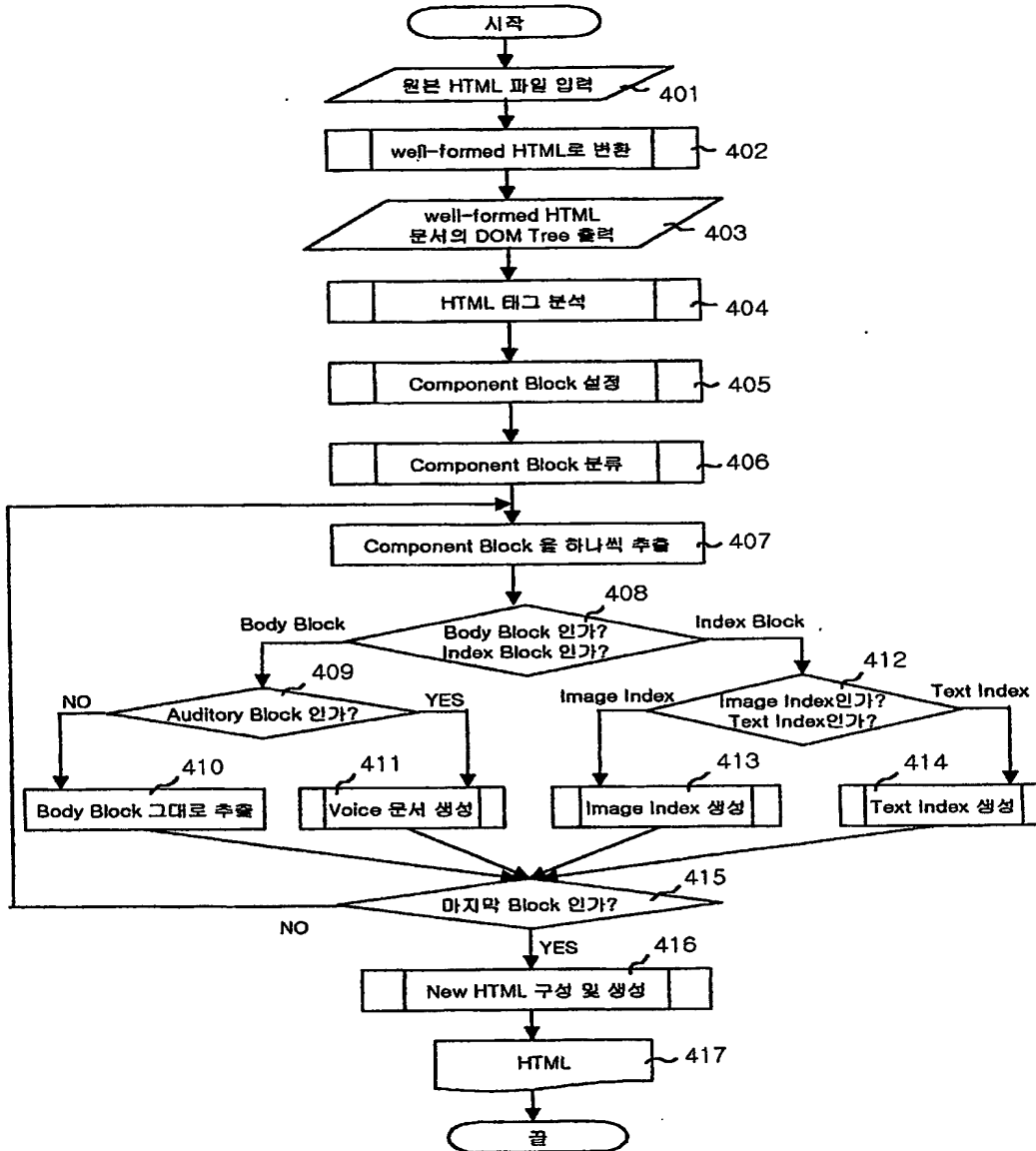
【도 2】



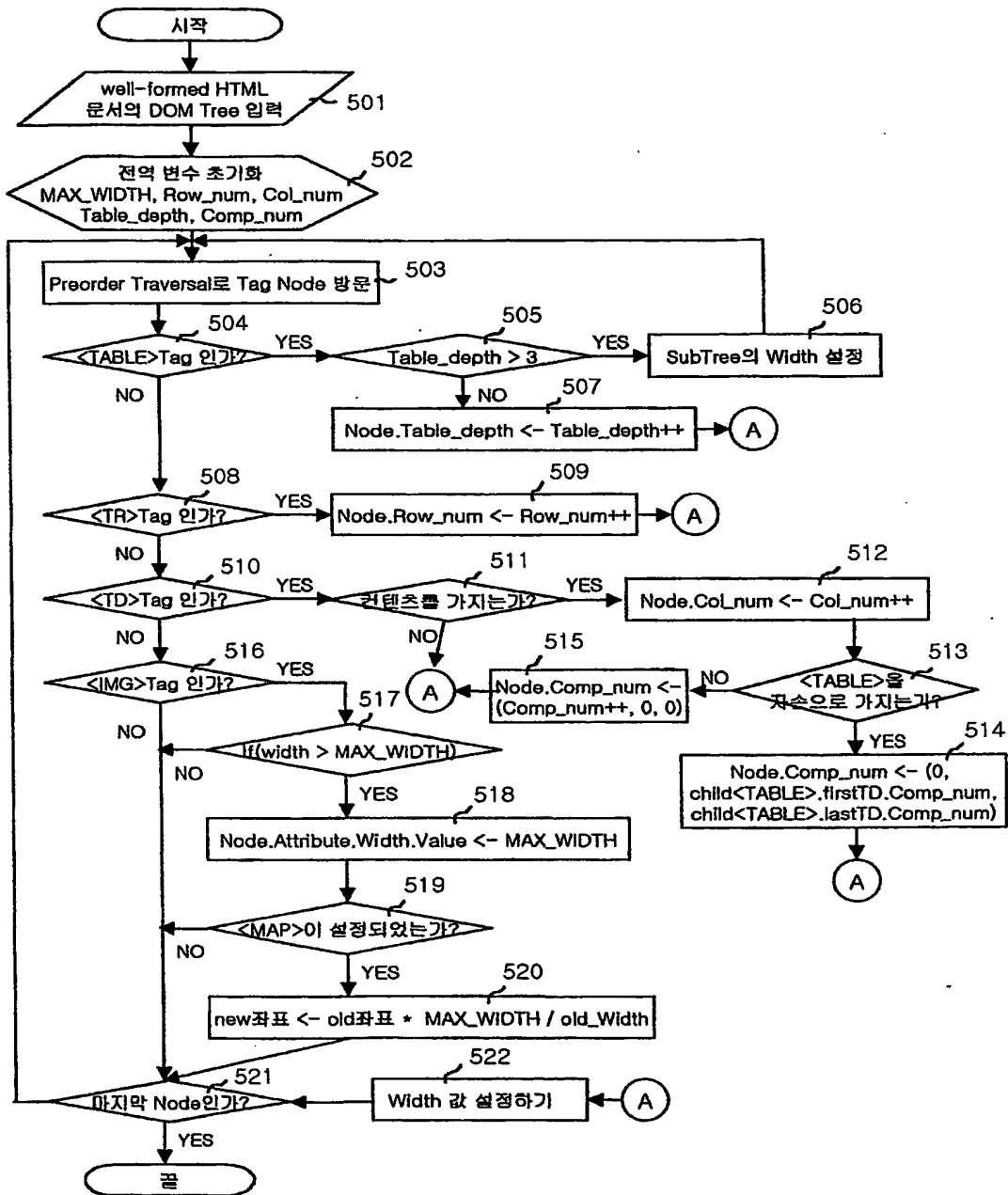
【도 3】



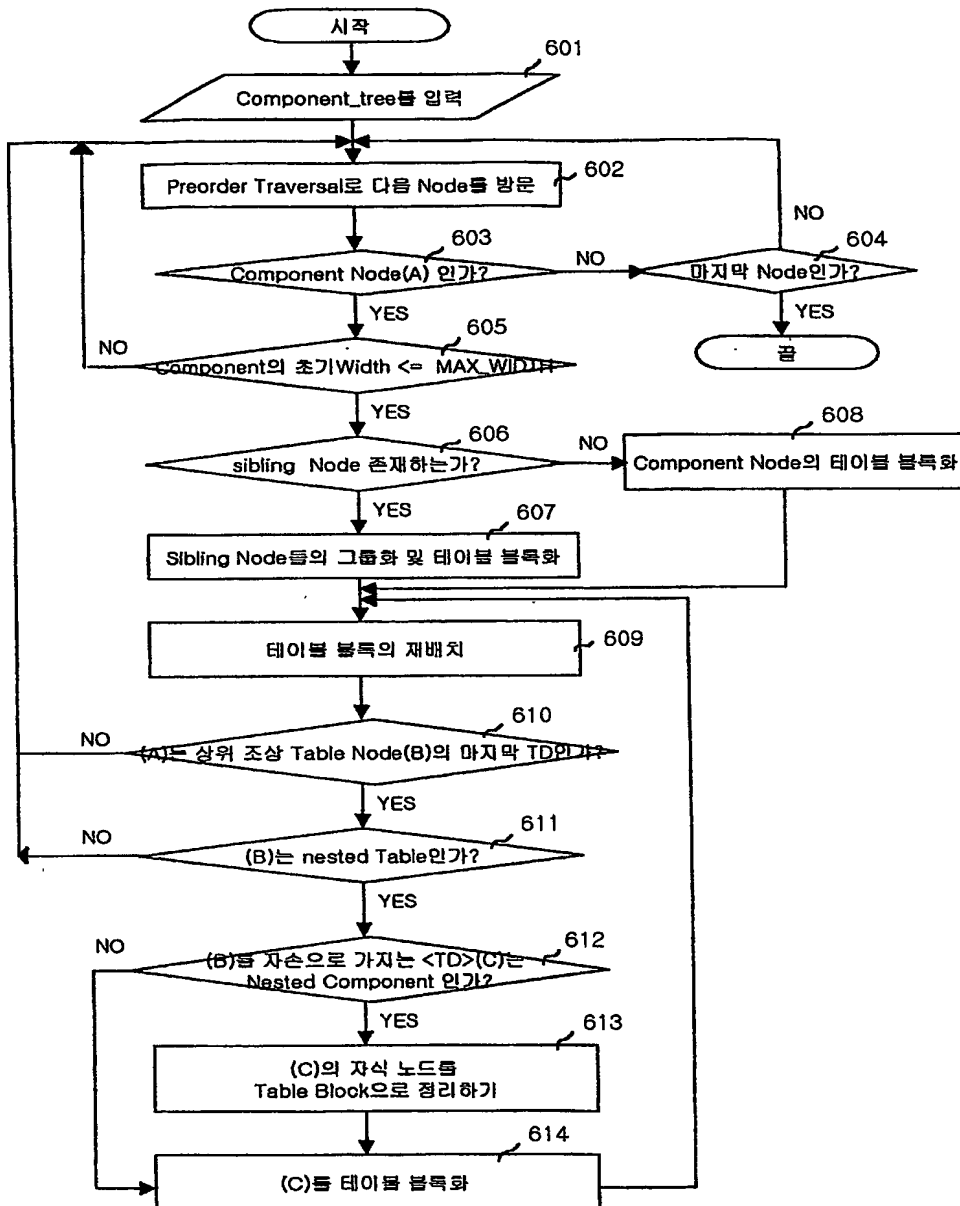
【도 4】



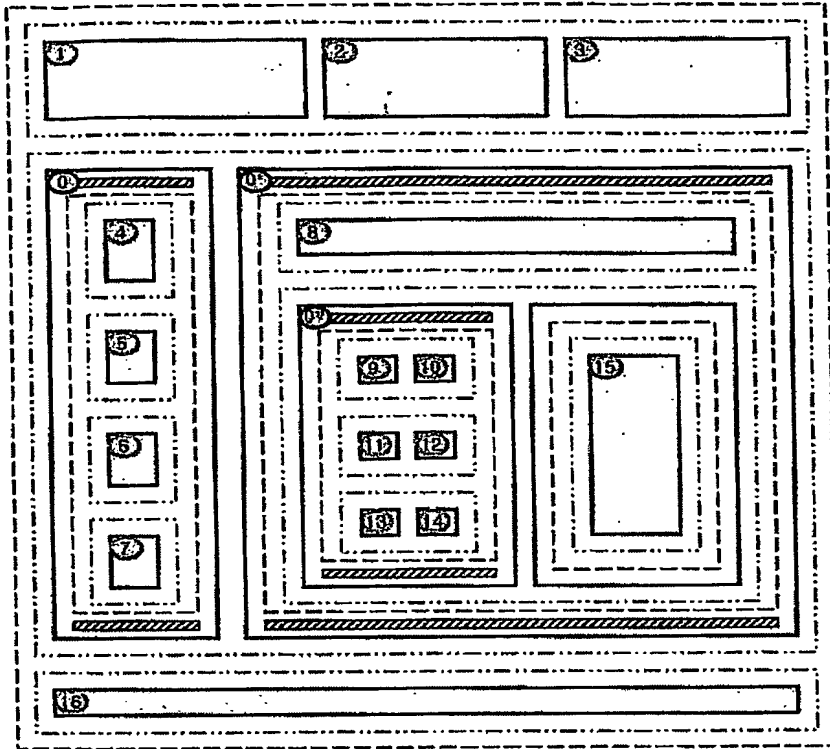
【도 5】



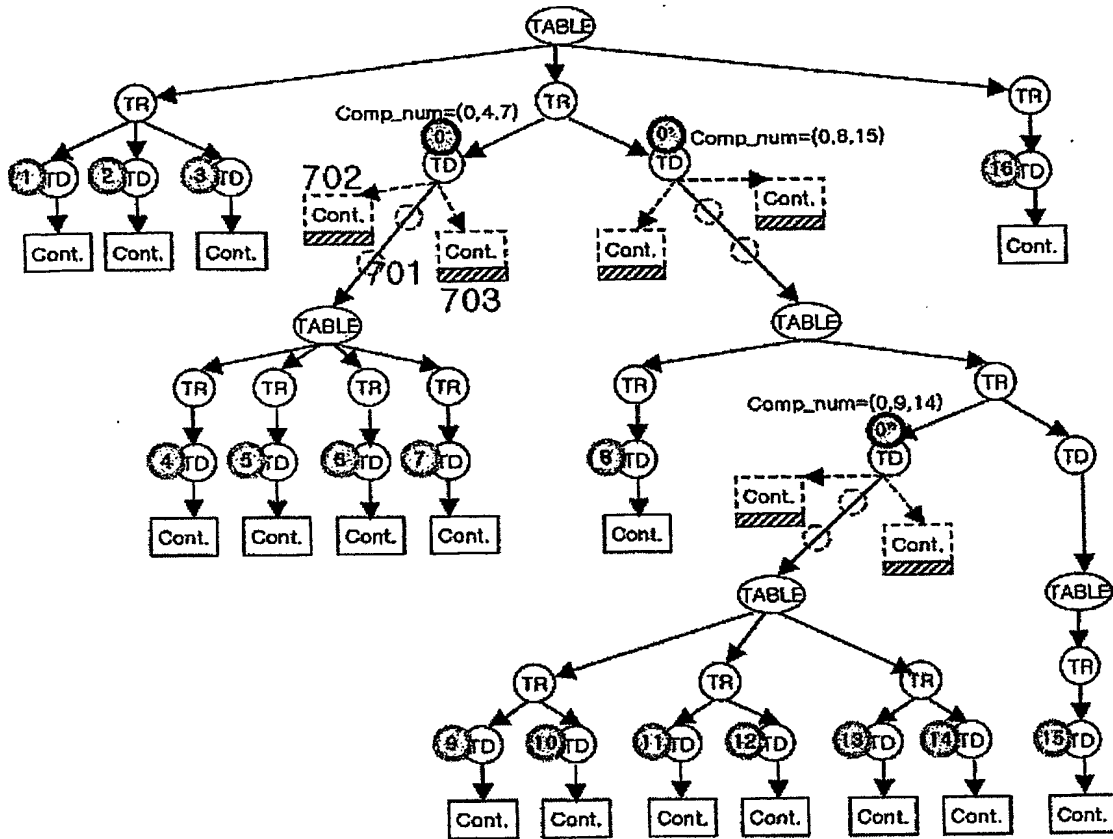
【도 6】



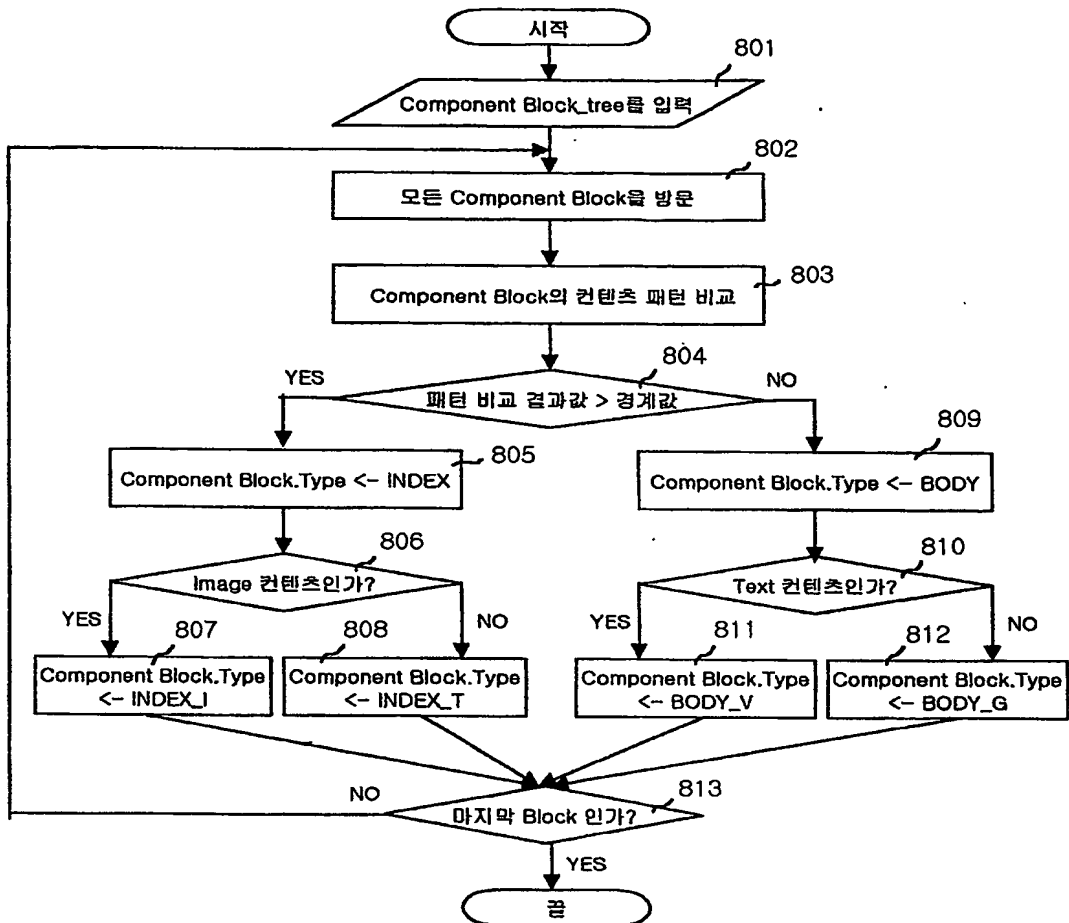
【도 7a】



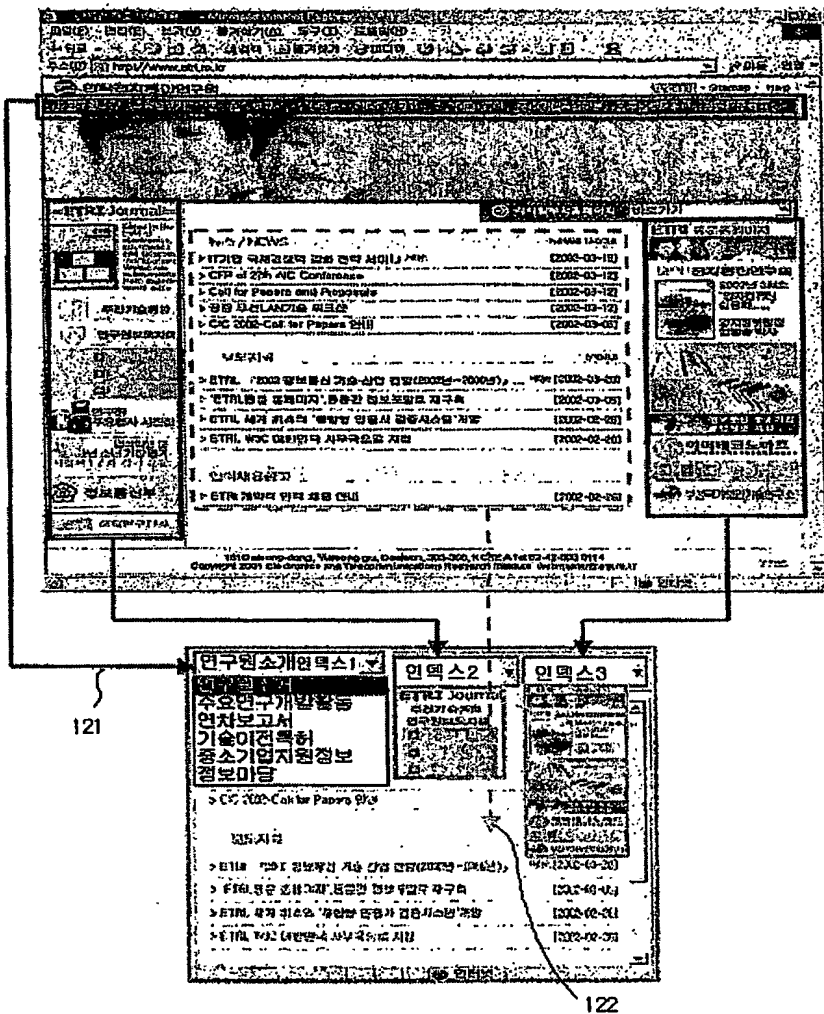
【도 7b】



【도 8】



【도 9a】



【도 9b】

